Transformations to obtain constant variance

Suppose a plot of e_i against $\hat{y}_i$ indicates that the variance is not constant. Transformations can help.

Let $z = g(y)$ be a transformation of y

A first order Taylor approximation gives

$$z \approx g(\mu) + g'(\mu)(y-\mu) \quad \text{where } \mu \text{ is some constant.}$$

Hence a random variable $Z = g(Y)$ should be possible to approximate by.

$$Z \approx g(\mu) + g'(\mu)(Y-\mu) \quad \text{where } \mu = E[Y]$$

Thereby $\quad Var(Z) \approx g'(\mu)^2 \, Var[Y]$. ~~For the variance of Z to be~~

~~≈ constant we must have~~ To have

$$g'(\mu)^2 \, Var[Y] = k. \quad \Rightarrow \quad \cancel{Var[Y] = \frac{k}{g'(\mu)^2}} \quad g'(\mu) = \frac{\sqrt{k}}{SD(Y)}$$

We get.

$SD(Y) \propto \sqrt{\mu} \quad (Var(Y) \propto \mu) \Rightarrow g'(\mu) = \dfrac{\sqrt{k}}{c\sqrt{\mu}}$ and an appropriate transformation will be $g(Y) = \sqrt{Y}$

$SD(Y) \propto \mu \quad (Var(Y) \propto \mu^2) \Rightarrow g'(\mu) = \dfrac{\sqrt{k}}{c\mu}$ and appropriate transformation $g(Y) = \ln(Y)$

$SD(Y) \propto \mu^2 \quad (Var(Y) \propto \mu^4) \Rightarrow g'(\mu) = \dfrac{\sqrt{k}}{c\mu^2}$ and appropriate transformation $g(Y) = \dfrac{1}{Y}$

Transformations can also sometimes be used to transfer nonlinear models into a linear one.

$$Y_i = \alpha e^{\beta x_i + \varepsilon_i} \Rightarrow \ln(Y_i) = \ln\alpha + \beta x_i + \varepsilon_i'$$

$$Y_i = \frac{x_i'}{\alpha + (\beta + \varepsilon_i') x_i'} \Rightarrow \frac{1}{Y_i'} = \frac{\alpha}{x_i'} + \beta + \varepsilon_i'$$

Box-Cox transformation

Suppose $Y_i > 0 \ \forall i$.

1. Consider a value of λ from a grid on a selected range $[(-2, 2), (-1, 1)]$

2. For each value of λ compute new responses

$$V = \begin{cases} \dfrac{Y^\lambda - 1}{\lambda \dot{Y}^{\lambda - 1}}, & \lambda \neq 0 \\ \dot{Y} \ln Y, & \lambda = 0 \end{cases}$$

where $\dot{Y} = \left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}}$ i.e. the geometric mean of the Y's

Hence $\underset{\sim}{Y} = (Y_1, Y_2, \dots, Y_n)^T \rightarrow \underset{\sim}{V} = (V_1, V_2, \dots, V_n)^T$

3. Regress $\underset{\sim}{V}$ on $\underset{\sim}{X}$ and find $SS_E(\lambda)$

4. Plot $SS_E(\lambda)$ versus λ for each value of λ. Draw a smooth curve and find $\lambda = \lambda^0$ that gives the smallest SS_E.

If λ^0 is close to -1, 0 or $1/2$ use these values instead for an easier interpretation.

5. Use the transformation Y^{λ^0}

Confidenceintervals for β_j, $j = 1, 2, \ldots, k$.

from

$$P\left(-t_{\frac{\alpha}{2}, n-k-1} \le \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{c_{jj}}} \le t_{\frac{\alpha}{2}, n-k-1}\right) = 1-\alpha$$

$$\Leftrightarrow P\left(\hat{\beta}_j - t_{\frac{\alpha}{2}, n-k-1}\hat{\sigma}\sqrt{c_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\frac{\alpha}{2}, n-k-1}\hat{\sigma}\sqrt{c_{jj}}\right) = 1-\alpha$$

Confidence ellipsoid for β

$$\varepsilon \sim N(\underline{0}, \sigma^2 I) \Rightarrow (\underline{\hat\beta} - \underline{\beta})^T \frac{\underline{X}^T \underline{X}}{\sigma^2} (\underline{\hat\beta} - \underline{\beta}) \sim \chi^2(k+1)$$

Now $\dfrac{SS_E}{\sigma^2} \sim \chi^2(n-k-1)$ and $\dfrac{SS_E}{n-k-1} = \hat\sigma^2$

Therefore $\dfrac{(n-k-1)\hat\sigma^2}{\sigma^2} \sim \chi^2(n-k-1)$

We get

$$\frac{\dfrac{(\underline{\hat\beta} - \underline{\beta})^T \underline{X}^T \underline{X} (\underline{\hat\beta} - \underline{\beta})}{\sigma^2 (k+1)}}{\dfrac{(n-k-1)\hat\sigma^2}{\sigma^2 (n-k-1)}} \sim F_{k+1,\, n-k-1}$$

and a $100(1-\alpha)\%$ confidence region for $\underline{\beta}$ is given by

$$(\underline{\hat\beta} - \underline{\beta})^T \underline{X}^T \underline{X} (\underline{\hat\beta} - \underline{\beta}) \le (k+1)\hat\sigma^2 \, f_{\alpha,\, k+1,\, n-k-1}$$

Confidence interval for expected response.

Construction of confidence intervals for the expected response given values $x_{10}, x_{20}, \dots, x_{k0}$ for $x_1, x_2, \dots, x_k$

Since $\hat\beta$ is independent of $\hat\sigma^2$, $\underline{\hat y} = \underline{X}\underline{\hat\beta}$ must also be independent of $\hat\sigma^2$.

Further $E[y \mid x_{10}, x_{20}, \dots, x_{k0}] = \beta_0 + \beta_1 x_{10} + \dots + \beta_k x_{k0}$

$$= [1, x_{10}, x_{20}, \ldots, x_{k0}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \underline{x}_0^T \underline{\beta} = \mu_{Y|x_{10}, \ldots, x_{k0}}$$

$$\hat{\mu}_{Y|x_{10}, \ldots, x_{k0}} = \underline{x}_0^T \hat{\underline{\beta}} = \hat{Y}_0$$

$$Var(\hat{Y}_0) = E[\underline{x}_0^T(\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})^T \underline{x}_0] = \underline{x}_0^T E[(\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})^T] \underline{x}_0$$

$$= \sigma^2 \underline{x}_0^T (\underline{X}^T \underline{X})^{-1} \underline{x}_0$$

Since $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are independent we get:

$$T = \frac{\hat{Y}_0 - \mu_{Y|x_{10}, \ldots, x_{k0}}}{\hat{\sigma}\sqrt{\underline{x}_0^T (\underline{X}^T \underline{X})^{-1} \underline{x}_0}} \sim t_{n-k-1}$$ and a $100(1-\alpha)\%$ confidence interval for expected response when $\underline{x} = \underline{x}_0$ is

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-k-1} \hat{\sigma} \sqrt{\underline{x}_0^T (\underline{X}^T \underline{X})^{-1} \underline{x}_0}$$

Prediction interval for a new value of the response when $\underline{x} = \underline{x}_0$

Let Y_0 be the new value then

$$E[Y_0 - \hat{Y}_0] = E[\underline{x}_0^T \underline{\beta} + \underline{\varepsilon} - \underline{x}_0^T \underline{\beta}] = \underline{x}_0^T \underline{\beta} - \underline{x}_0^T \underline{\beta} = \underline{0}$$

$$Var[Y_0 - \hat{Y}_0] = Var[Y_0] + Var[\hat{Y}_0] = \sigma^2 + \sigma^2 \underline{x}_0^T (\underline{X}^T \underline{X})^{-1} \underline{x}_0$$

such that $$T = \frac{\hat{Y}_0 - Y_0}{\hat{\sigma}\sqrt{1 + \underline{x}_0^T (\underline{X}^T \underline{X})^{-1} \underline{x}_0}} \sim t_{n-k-1}$$ and a

$100(1-\alpha)\%$ prediction interval for Y given $\underline{x} = \underline{x}_0$

is $$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-k-1} \hat{\sigma} \sqrt{1 + \underline{x}_0^T (\underline{X}^T \underline{X})^{-1} \underline{x}_0}$$